IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR PATENT

ADDING IMPERCEPTIBLE NOISE TO AUDIO SIGNALS TO CAUSE SIGNIFICANT DEGRADATION WHEN COMPRESSED AND DECOMPRESSED

Inventors: Paul R. Goldberg
Vlad Fruchter

CROSS-REFERENCE TO RELATED APPLICATIONS

This is a continuation-in-part of co-pending patent application serial no. 09/570,655, filed May 15, 2000. This is also related to patent applications of Paul R. Goldberg, serial no. 09/484,851, filed January 18, 2000, and its continuation-in-part application serial no. 09/584,134, filed May 31, 2000. (Hereinafter referred to as the "Secure Transmission Patent Applications.") These three applications are expressly incorporated herein by this reference.

BACKGROUND OF THE INVENTION

This invention is related to the processing, transmission and recording of audio signals such as music, and, more specifically, to techniques that prevent or discourage the unauthorized copying and/or distribution of audio content.

The ease that music can be electronically distributed by private individuals over the Internet has caused great concern on the part of the music content providers, their distributors and retailers. It is now possible for one compact disc to be purchased and, in a matter of hours, electronically distributed by the purchaser without charge to his or her friends, and even to people or enterprises unknown to the purchaser. Clearly, this reduces the desire of many to pay for the music and causes great concern on the part of the recording industry that their revenues and profits will be significantly eroded. If this should happen, record labels may react by no longer promoting and offering the wide range of artists and music

styles they currently do. Only those music products that can provide a rapid and sure return on investment will be supported. The result will be that the music buying public's choice will be limited. Many of the artists and types of music we have come to enjoy will no longer be available.

What makes this unprecedented electronic sharing of music over the Internet practical is the availability of high caliber audio compression algorithms. These algorithms are capable of reducing the data rates and data volumes, previously required to digitally represent music, by a factor of more than 10, while maintaining acceptable audio quality. The provider compresses the music data by such a factor and the recipient then applies a mating decompression algorithm to the received compressed data to recover something close to the original music. MP3 (MPEG 1 Layer 3) and AAC (Advanced Audio Coding) are examples of commonly used compression algorithms that offer this capability. DTS (Digital Theater Systems) and AC-3 compression algorithms are professionally used for movie sound tracks and the like. A common characteristic of these compression algorithms is that data of frequencies not separately resolvable by the human ear are discarded, thereby to reduce the amount of data necessary to be transmitted.

Psycho acoustic audio compression technologies, such as MP3 and AAC, operate by making quantized noise imperceptible to the human hearing system. In digital audio systems, such as those used by compact disks to deliver music to consumers, 16 bit resolution is considered to be about the practical minimum number of bits to use to keep the quantized noise down to an acceptable level (in this case about 96dB below the maximum signal level). The objective of an audio compression algorithm is to use as few a bits as possible to represent the input audio signal. In order to use fewer bits, mechanisms need to be found to minimize the increased level of quantized noise, or make this higher level of noise indiscernible to the listener. The characteristics of the human hearing process provides several opportunities to do the latter. The first is the basic threshold of hearing. Human ears tend to be less sensitive at low and high frequencies. The second characteristic can be seen by considering the structure of the inner ear. The cochlea is a spiral,

tapering passage with the basilar membrane that is stretched, more or less, across the diameter along its length. Sound is conducted from the outer ear to the fluid in the cochlea where it travels the length of the basilar membrane. Different frequency components of a sound vibrate the hair cells at different locations along the membrane, stimulating the auditory nerves. The frequency dependent movement of the hair cells make the ear act like a spectrum analyzer. A high level frequency component will not only vibrate the hair cells at the location sensitive to that specific frequency, but it will also vibrate the hair cells at some of the adjacent locations as well. The spreading of the response to a specific frequency over multiple hair cell sensors can and will override, or "mask", the response to other lower level, nearby frequency components. The ability of relatively loud sounds to mask lower level ones is usually described by sets of frequency and level-dependent "masking curves". If the quantizing noise produced by a coarse quantizer can be confined to the spectral region near to the signal component being quantized (or encoded), and if that noise is low enough to fall below the masking curve of the signal being coded, then the listener will not hear the quantized noise. That is, the amount of data that represent spectral regions near to the signal component being quantized can be reduced without it becoming noticeable to the listener.

What is needed is a means to permit this technology to serve the recording industry's need for revenue and profits, by allowing Electronic Music Distribution ("EMD") to be used as another channel of distributing and collecting revenue for music product, while simultaneously preventing this same technology from negatively impacting the industry. The present invention is directed in large part to satisfying this need.

SUMMARY OF THE INVENTION

Briefly and generally, an audio signal is modified in a manner that is not perceptible to the human ear until, after compression according to one of various specific compression algorithms, an uncompressed signal is noticeably distorted to the human ear. In a first embodiment, the original audio signal is so modified, so

that any such compression and decompression results in the distorted signal. In a second embodiment, a compressed audio signal is provided that gives a high quality signal when decompressed but which, when that decompressed signal is again compressed, its further decompression results in a noticeably distorted signal. The effect of providing a sound signal that cannot be compressed without such degradation of quality limits its distribution over the Internet since it is not currently practical to distribute uncompressed sound signal files over the Internet. The time taken to transmit uncompressed files and the computer storage space necessary to hold them are far too large for the usual Internet user. Therefore, illegal distribution of music over the Internet will be significantly reduced. Sales by music providers will be maintained.

In a first example of the first embodiment of the present invention, an audio signal is modified by increasing levels of its masked frequency components while still retaining those levels below the masking level of a typical human ear. The resulting distortion caused by this "anti-compression" processing of the signal is thus not heard by a listener. But when the modified audio signal is compressed and then decompressed by algorithms of the type discussed above, the resulting sound is significantly degraded in quality. This is because the compression algorithm is operating on a different sound signal than the original one that is desired to be reproduced. As a result, the masking levels are different and the reduced number of bits used to represent the spectrum are thus allocated differently. When these different bit allocations are used to reconstruct the sound signal, it does not represent the original signal. Indeed, the compression algorithm may need to allocate a limited number of bits to an expanded portion of the signal's spectrum, thus not representing the unmasked, audible portions with enough resolution. The resulting decompressed sound signal is a significantly degraded version of the original signal and is therefore not desirable for listening.

In a second example of the first embodiment of anti-compression, relationships between multiple audio data channels are used. The example of this embodiment employs the alteration of timing and or phase relationships between

such channels. Alteration of these relationships causes subsequent compression and decompression processes to incorrectly combine multiple channel data, and thus cause a degraded version of the original audio signal to be produced.

A third example of the first anti-compression embodiment also uses the relationships between multiple audio data channels, but in this case they are used to unmask data embedded into the original signal that are masked by the audio data prior to the compression process being performed.

In a second embodiment of the present invention, an encode/decode compression algorithm pair is described which has the characteristic of producing compressed audio data that can be decompressed for listening, but cannot be compressed with quality for a second time, thus effectively disallowing retransmission of the audio data over the Internet. A first example of the "one generation" codec with built in anti-compression processing, uses the addition of noise or other data to achieve the desired unique results.

A second example of the second embodiment employs the generational characteristics of compression algorithms to a similar end.

A third example of the one generation codec embodiment of the present invention uses the fact that compression algorithms with improved generational qualities often use additional techniques to reduce bit requirements without adding quantization noise. These techniques form the basis of additional methods for producing compressed audio data that can be decompressed for listening, but cannot be compressed with quality for a second time.

In a fourth example of the one generation codec embodiment of the present invention, an alteration of the timing of the processing of defined blocks of audio data is employed to create a compressed version of the original audio data that displays high quality when decompressed and listened to, but will cause following compression and decompression processes to be unable to choose the size and process timing necessary to mask, transient noise added to the audio data during the initial compression process.

Finally, a unique method of adaptively optimizing anti-compression processing of audio data is also included as part of the present invention. For example, any of the foregoing processing techniques can be adjusted as a function of characteristics of the input audio signal being processed during such processing.

Rather than using the principles underlying compression algorithms to reduce the amount of audio signal data while maintaining quality, the techniques of the present invention apply those principles to change the character of the sound signal so that it cannot be compressed without significant degradation in the quality of the signal. Indeed, existing compression algorithms have been designed to allow a signal to be compressed and decompressed two or more times without significant degradation of the quality of the signal that is perceptible to the human ear, termed their "generational" quality. But the present invention uses the principles of compression in a reverse manner, modifying a sound signal so that it will not retain its quality when compressed. This contrary use of the principles underlying compression algorithms greatly improves the ability of a music provider to control the distribution of its music.

Additional features, advantages and objects of the present invention are included in the following description of its embodiments, which description should be taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the processing of an audio signal according to the present invention;

Figure 2 is a curve representing an audio signal being processed;

Figure 3 is an example frequency spectra for a block of the audio signal that shows its processing according to the present invention;

Figure 4 shows an example frequency spectra for a block of the audio signal after it is modified by the processing of the present invention;

Figure 5 illustrates a recording application of the present invention;

Figure 6 illustrates an Internet music delivery application of the present invention;

Figure 7 shows a key card for use in the delivery application of Figure 6;

Figure 8 illustrates a one generation codec with built-in anti-compression components as part of the compression process; and

Figure 9 illustrates the application of "adaptive processing", referred to as optimization, to maximize the difference between the high quality of a processed but not compressed audio signal as compared with the reduced quality of a processed and compressed audio signal.

DESCRIPTION OF EXEMPLARY EMBODIMENTS

The block diagram of Figure 1 shows an example anti-compression signal modification system 511 of the first embodiment of the present invention, which operates to process an input audio signal 513. The first three processing steps 515, 517 and 519 are substantially the same as those of a compression algorithm of the type discussed above. In the step 515, a block of data of the signal 513 is acquired. Referring to Figure 2, a portion 527 of the signal is shown divided into time successive blocks, such as blocks 529 and 531. Preferably in a digital format, data representing samples of the signal 527 during a block are quantized in the step 515. The signal block is then filtered in a step 517 in order to obtain floating point coefficients of the frequency spectrum of the block of data. Each sampled frequency is expressed as an exponent (coarse measure) and mantissa (fine). Those values are then used by a non-linear quantizer 519 to calculate a masking function 535 (Figure 3) and compare it to the spectrum 533 of the block. When used as part of a compression algorithm, the quantizer 519 also allocates a lesser number of bits than in the incoming signal 513 to represent the signal in limited frequency ranges 537 where the spectrum 533 is greater than the mask 535. The remaining frequency ranges are not necessary to be included in the compressed signal since they are below the levels indicated by the mask 535 that a human ear can hear. So they can

be omitted, and it is this omission that allows the amount of data representing the signal to be reduced.

But since, in the technique being described, the input signal is not being compressed, the bit allocations for the limited frequency ranges 537 need not be calculated. Rather, a step 521 is added that does not exist in compression algorithms. This step calculates increases that can be made to various frequency components of the incoming signal 513. The block spectrum 533 and mask 535 calculated in the non-linear quantizer 519 are used in this calculation. This calculation increases the value of frequency components that are less than the mask 535, increasing the signal spectrum 533 into shaded regions 539 of Figure 3. Since, as expressed by the masking function, the human ear cannot separately resolve these frequencies, this will not be perceived to degrade the signal, so long as the spectrum 533 is not increased above the level of the mask 535. Indeed, it is preferable to maintain the spectrum 533 below the mask 535 by some margin in the regions 539 to assure that these added signal components will not be heard by the human ear. Example margins are ten or twenty percent of the level of the masking function 535.

Furthermore, all frequencies in the regions 539 need not be raised above the levels of the curve 533. The spectrum 533 needs to be altered only enough to result in a subsequent application of a compression and decompression algorithm to the modified signal to cause undesirable perceptible distortions of the original signal 513.

And, as a further feature, the level of some frequency components of the signal 533 may be increased above the mask 535 without affecting the quality of the sound to the human ear, such as at frequencies adjacent peak frequency levels of the spectrum. This type of change to the signal 533 can also affect the ability of a decompression algorithm operating on a compressed version of the altered signal to provide a good quality decompressed signal.

Alternatively, changes to the spectrum 533 may be more modest so that the modified signal can be subject to one compression and decompression cycle without significantly degrading the quality of the incoming signal 513 but would

result in serious degradation if again compressed and decompressed. This partial degradation has application to the Internet, wherein the partially degraded signal is initially sent over the Internet and re-transmissions of the audio signal are discouraged when the second or more cycle of compression and decompression makes the sound undesirable. This application is discussed below with respect to Figure 8.

In any event, the additional calculated signal is then added to the input signal 513 at 523 in order to provide a modified signal output 525. An implementation of the processing of Figure 1 includes a digital signal processor that operates under controlling software to perform the functions described above.

The step 521 may determine in one of several ways the amount that the level of the audio signal 513 is to be increased in the step 523 over a portion or all of the frequency ranges 531. One way is to generate random or pseudo-random noise that is uncorrelated with the signal 513 and add appropriate levels of such noise to the signal in the block 523. Another way is to generate a defined signal, such as a sine wave or a combination of sine waves of different frequencies, that is uncorrelated with the audio signal, and then add such a signal(s) to the audio signal.

A further way to modify the audio signal 513 is to add an amount of signal data that is correlated to it. This last technique may be implemented by simply increasing the levels of the frequency components already in the signal that are below the masking curve 535. This preserves the original audio qualities of the initial signal because the added data is correlated with that signal. The added data is then also difficult to distinguish from the original signal when listening to the resulting output audio signal 525. One way to increase the signal levels is to multiply the levels of some or all of the various frequency components of the audio signal 513 within the frequency ranges 539 by a frequency dependent factor greater than unity to increase the level of some or all of such frequencies to a level that is equal to or some defined amount below the masking function 535.

Yet another way to modify the audio signal of 513 is to add a replica of the original signal from one or more frequency bands, position shifted in time by

one or more clock cycles with respect to the original audio signal, to the original audio signal. The original audio qualities of the initial signal are preserved because the added data is presented in very rapid sequence with respect to the original data and is correlated with the original audio signal. Here again, the added data is also difficult to distinguish from the original signal when listening to the resulting processed output audio signal 525. One way to add this replicated time shifted data is to store a block of the original audio signal's frequency domain coefficients, delay this coefficient data in time, recreate a time domain representation from the frequency coefficient data, and add this delayed time domain data back to the time domain representation of the original signal. Another way is to first use a narrow band filter bank in the time domain to separate the frequency components of the original signal into multiple narrow bands. Then select which frequency band or bands of the original audio data are most beneficial to replicate and delay by one or more clock cycles with respect to the original audio data, based on which one of these frequency components will require the most bits to accurately represent the original signal in a compressed version of the original signal. Then amplitude normalize these frequency components with respect to the original signal, such that their amplitude is above, equal to or below the masking curve amplitude defined by the frequency components of the original audio signal, based on the masking properties associated with each band of frequencies. Then time synchronize this frequency band data, and combine it with the original audio data. Subsequent compression of an audio signal processed in either of these manners is degraded because a compression algorithm will allocate additional bits to the added time shifted data in an effort to maintain the quality of the compressed audio.

The curves of Figure 4 illustrate the effect of one specific application of the signal processing described with respect to Figures 1-3. A frequency spectrum 541 is shown for a block of the output audio signal 525 in the same time interval as illustrated in Figure 3. The input signal 513 has been modified by increasing the level of the spectrum 533 in all frequency ranges where it was below the mask 535 (shaded regions 539) up to the level of the mask 535. This

represents the maximum increase of the input signal 513 that is desirable, and, as discussed above, is normally more than what is normally prudent to add. The main point to note from Figure 4 is that the output signal 525 now has a different frequency spectrum than the input signal 513. If the output signal is then compressed by the type of algorithm discussed above, a resulting mask 543 is different. The mask of a block is calculated as part of compression algorithms from the frequency spectrum of the block itself and, in some algorithms, from data of the frequency spectra of adjacent blocks occurring in time before and/or after the block represented by Figure 4.

The example shown in Figure 4 shows a large extent 545 of frequencies where the spectrum 541 is higher than the mask 543. The compression algorithm then must allocate its limited number of bits across the frequency bands 545 which are much larger in extent of frequency than the bands 537 (Figure 3) of frequencies for the original signal 513. Further, the signal spectrum 541 (Figure 4) of the output signal 525 is much different than the spectrum 533 (Figure 3) of the input signal 513, differences being noted over ranges 547 of frequencies. At the same time, the increased signal has the effect of causing the signal spectrum 541 and the mask 543 calculated (at least in part) from it to follow each other more closely (curves of Figure 4 vs. those of Figure 3). This also makes the signal less compressible after the signal has been increased. The result is a compressed signal calculated from the output signal 525 that is much different than one calculated from the input signal 513. The output signal 525, because of the nature of the data intentionally added to the input signal 513, does not lend itself to compression if a faithful reproduction of the input signal 513 is desired upon decompression.

Like psycho acoustic based compression processes, the embodiment described above transforms the complex audio signals that are input to the system into the frequency domain, and masking curves for the different signal components are computed. The masking (hearing) threshold curves are compared with the spectrum of the input audio signal, and the limits on the level of quantizing noise or other added data that can be "hidden" by the audio signal input to the system is thus

determined. In the compression processing case, the encoder then makes decisions about the coarseness of the quantizer, or the number of bits that need to be assigned to each of the frequency components of the audio signal, in order to assure that the added quantizing noise, caused by the coarser quantizing process, is masked and thus imperceptible to the listener. In the case of the techniques being described herein, however, this information is employed to determine how much extra noise, for example, can be added to the original audio signal input to the system, before this noise can be heard by the listener. Unlike the compression processing case, in which the output signal is the lower data rate, more coarsely quantized signal, the present techniques output the original signal with noise added on a frequency component by frequency component basis, the level of added noise chosen to be just low enough to be masked by adjacent frequency components in the original audio signal. The audio output signal then no longer has the uniform low level noise floor of the original input audio signal. Instead it has a dynamically changing, program dependent noise floor. If this digital audio signal is converted into its analog audio presentation and listened to, the added noise will properly be masked by the adjacent higher level frequency components in the signal, and thus not heard. If, however, this processed signal is fed into a compression encoder/decode process for Internet distribution, the additional quantizing noise caused by this following audio compression/decompression process will add to the noise injected into the audio signal by the techniques described above. The resulting audio signal will then contain a total noise which is over the masking curve limit, and thus the noise will be perceptible to the listener. These noise artifacts will make the compressed audio signal unsuitable for distribution over the Internet, which is an objective of the present invention. It should be noted that the injected "noise" can have a wide range of characteristics. These characteristics are chosen to be most annoying to the listener in the event the noise is made perceptible by a follow-on compression process.

In another approach to modifying the audio signal 513, timing and or phase relationships between channels of a dual channel (stereo) audio signal are

modified. This added data can be a fixed phase or timing change or one that changes over time. In addition, the altered phase or timing relationship can be different for each audio frequency encountered in the original audio signal. "Intensity" and "joint" stereo compression processes, which are well know in the art, employ the technique of combing channel data above a certain frequency and/or employing a common bit pool. The use of a common bit pool permits the assignment of bits to channels depending on the data rate requirements of the audio data appearing in each channel. By altering the phase or timing of the information in these channels with respect to each other, common data appearing in both channels cancel or partially cancel during the compression process. This results in an output after the decompression processes which varies in amplitude, quite unlike the original stereo audio signal. By this means a degraded version of the original audio signal will be produced after the compression/decompression cycle, but, because human hearing cannot easily detect phase variations, the stereo audio will sound normal before the compression/decompression process.

A simple implementation of the above concept calls for advancing or retarding the phase by a predetermined number of degrees, for example 180 degrees, of all frequencies above a predetermined frequency, for example 1500 Hz, in one channel of a stereo audio signal with respect to the second channel of the stereo audio signal. This process produces an audio signal which sounds identical to the original stereo audio signal, but will be degraded by a subsequent compression process which employs joint or intensity stereo techniques. The resulting joint stereo compressed and decompressed audio signal sounds very much as if it is emanating from an underwater source because of the amplitude variations introduced in the audio program material by complete or partial phase cancellation as described above. A similar effect can be produced if instead of introducing 180 degree phase inversion above a predefined frequency, one of the two stereo audio channels is advanced or retarded in time with respect to the other channel. This can be implemented in the digital domain by advancing or retarding one of these two channels with respect to the other channel by 1 or more bits.

A more advanced version of the above concept calls for modulating the timing and or phase of a particular frequency or frequencies. For example a rate below or above the lowest or highest frequency the human ear can detect can be employed. Such a rate could be 1 Hz. The modulation would be imposed on one or more frequency component present in one channel of a stereo channel pair as compared to the other channel of the stereo channel pair. This phase modulation cannot be heard in the processed original stereo audio data, but when compressed and decompressed, using an intensity stereo compression algorithm for example, would produce an output which is quite degraded. A compression algorithm which employs intensity stereo techniques combines input audio data above a predefined frequency, and retains only the intensity of the total energy appearing at each frequency band above this predefined frequency. The varying amplitude introduced in the common data present in the channels, due to the modulation of energy in each channel which are at like frequencies but at varying phases with respect to each other, produces audio which fluctuates in amplitude and is quite unpleasant to listen to.

An additional method of modifying audio signal 513 also uses the relationships between multiple audio data channels. In this case spurious data which is masked by the original audio signal is embedded into each channel of the original audio signal. This data is caused to be"unmasked" when the audio signal is compressed. One example of this approach is to first alter or totally reverse the phase of one channel of a stereo audio signal with respect to its other channel. This alteration in phase, which could be either fixed or varying in time, could be implemented on frequencies which lie above a predetermined frequency, over a range of frequencies, or over one or more bands of frequencies. The spurious data is then added in phase into both channels. By choosing the spurious data such that it is below the masking threshold of the original audio signal, the spurious data will be inaudible when this now processed audio signal is reproduced for listening. However, if this signal is compressed, using a joint stereo encoder, and then reproduced for listening, the original stereo audio signal will be reduced in amplitude

due to phase cancellation between the channels, while the spurious data will be increased in amplitude, due to phase addition. This will result in a reduced masking level and an increased spurious data level. It will then follow that the embedded spurious data will be above the lowered masking threshold and be audible to the listener.

A modification of the above strategy is to add spurious data at a selected frequency or frequencies, and times, to one channel of a stereo audio signal, phase shift this added data by 180 degrees, and add it to the second channel of the stereo audio signal. The intensity and frequency components of this added signal energy would be chosen to be below the masking threshold set by the audio data in each channel. Being 180 degrees out of phase the spurious data added to the two channels would additionally tend to cancel when reproduced either in free air, through speakers or through headphones, and thus be virtually inaudible to the listener. When the audio processed in this manner is encoded with a compression algorithm that sums the absolute values of one or more of the frequency components in each channel of said two channel audio signal in order to reduce the data rate requirements of the compressed signal, the absolute values of the embedded spurious signals in each channel will constructively add and the embedded spurious signals will become audible to the listener.

One important application of the signal modification system 511 described above is illustrated in Figure 5. After the music or other program material for reproduction on a Compact Disc ("CD") is assembled as a digital file, indicated by a block 551, that file is processed by one or more of the techniques described above to add signal data to the audio signals of the file before making a CD master recording 553 from it. The content of the resulting replica CDs that are sold to consumers cannot then be compressed without a significant loss of quality of the content signals when decompressed. The same techniques can also be used when storing or distributing audio content by other means such as with audio tape, as a component of a Digital Video Disc ("DVD"), or as the digital or analog sound track on a motion picture release print. Since such compression is currently required

before the audio content can be stored or distributed in several ways, such as storing in non-volatile semiconductor memory cards or transmission over the Internet or other communications network, unauthorized copying and distribution of the content is thus greatly discouraged. The degraded music or other audio content is of little value.

The block diagram of Figure 6 illustrates a use of the present invention in the distribution of music or other audio content over the Internet in a manner that greatly discourages copying and re-distribution of the content by the recipient over the Internet. A master audio source file 555 is compressed, as indicated by a block 557, and then encoded, as indicated by a block 559, in order to provide a secure transmission that can be decoded only by the intended recipient. The compressed and encoded digital signal is then transmitted over the Internet 561 to the intended recipient who, in the normal case, has paid the content provider for it. The recipient must then decode the incoming signal, as indicated by a block 565, by use of a key or other accepted technique, and then decompress it, as indicated by a block 567. At this point, however, the master audio source file 555 is available to the recipient in a decoded and decompressed form that can easily be distributed to others over the Internet by a recipient who is willing to violate the copyright of the content provider. But since such unauthorized distribution is practical only if the content file is first again compressed by the recipient, noise or other data is added to the decoded and decompressed content file by the recipient's audio player or other utilization device, as indicated by a block 569. The recipient can, however, reproduce the audio content without degradation after the audio signal has been modified. The content, in the form of an analog or pulse code modulated ("PCM") signal, for example, is applied to standard audio circuits 571 that drive a loud speaker or head phones.

Such a signal addition in the recipient's utilization device is made effective when the recipient has no effective choice but to receive an output of the content from his or her utilization device after the audio signal has been modified. In order to prevent the recipient from accessing the content signal before the signal

is modified in the step 569, the signal modification is preferably performed in a physically sealed module 115' that also includes the decoding function 565. A key necessary for decoding the signal is included within the module in a manner that renders it inaccessible to the recipient. Since the content provider can make it a condition of supplying the music or other content that the recipient use such a sealed module to decode the transmitted encoded content, the added security against the recipient being able to easily redistribute the audio content is conveniently included in the same sealed module. As can be seen from Figure 6, a decoded digital signal of the content is not available except within the sealed module 115'. An input to that module is an encoded signal which the recipient cannot decode except with use of the module. An output of the module 115' presents the content in a standard format, such as an analog or PCM signal, which could normally be re-digitized or otherwise manipulated by the recipient for unauthorized redistribution. But since such redistribution normally requires that the signal be compressed prior to doing so, the noise or other data that is added to the output signal by the processing step 569 makes that highly undesirable or even impossible.

The sealed module 115' is a variation of the module 115 described in the aforementioned Secure Transmission Patent Application, with a specific version shown in Figure 7 hereof, where the reference numbers are the same as used in the Secure Transmission Patent Application but with a prime (') added for corresponding elements that are modified herein. The primary, and perhaps only, component of the sealed module115' is a digital signal processor ("DSP") integrated circuit chip 135'. The primary difference here is the inclusion of signal modification software 573 in its non-volatile memory 147' in a manner that the user cannot access that software or defeat its use to add the anti-compression noise or other data before an audio signal is made accessible to the user (recipient) at an output of the module.

As described in the Secure Transmission Patent Applications, the module 115' is preferably implemented in the form of a small key card that is made personal to a particular user by storing decryption (decoding) key(s) in its memory 147' that are unique to the user. The key card is removably inserted into the user's

audio player when connected to the Internet, a kiosk in a music store, or other content providing device, in order to purchase content from a provider with use of the user's key(s) stored within the card. The key card is also inserted into the recipient's player, as well as others, in order to allow the received content to be played by the recipient while restricting the extent to which the content can be transferred to or played by others. By the controlled addition of noise or other data to the content signal output of the sealed key card, according to the techniques described herein, unauthorized distribution and use are further technically restricted.

Figure 8 shows a second embodiment of the present invention. In this second embodiment an encode/decode compression algorithm pair is described which has the characteristic of producing compressed audio data that can be decompressed for listening, but cannot be compressed with quality for a second time, thus effectively disallowing retransmission of the audio data over the Internet. A compression algorithm with this characteristic is called a "one generation" algorithm. The use of a one generation algorithm serves as an alternative to including anti-compression signal modification in the recipient's player, as described with respect to Figure 6 and 7. As depicted in Figure 8, an audio source file 577 is compressed with an available algorithm, as indicated by a block 579, and some noise or other data for the same purpose is added, as shown by a block 581. The amount that the audio signal is increased by 581 is below that which significantly affects the quality of the content when decompressed by the user. But it is sufficient to cause the quality of the content signal to be significantly degraded if the decompressed signal is again compressed with the type of algorithm described previously. In either of the versions of the first embodiment shown in Figures 6 and 7 or that of the second embodiment shown in Figure 8, electronic distribution of music or other content is facilitated. It should be noted that the block 581 can be combined with the block 579 to form a single stage compression algorithm which provides a compressed audio output with anti-compression signal components added. In this case, a "calculate signal increases" block, such as block 521 of Figure 1, and an "adder" block such as block 525 of Figure 1, would be incorporated into the compression

algorithm itself, following the compression algorithm's non-linear quantizer block and preceding the compressed audio output from the compression algorithm.

A second approach applicable to the one generation codec embodiment described above employs the fact that compression algorithms inherently add quantization noise to the original signal during the compression process itself. As previous described, this is due to the fact that individual frequency components of the signal are more coarsely digitized in an effort to reduce the number of bits used to described the signal. This leads to "generation loss" when "cascading" compression processes. When compression algorithms are cascaded, that is a signal is compressed, then decompressed and then compressed and decompressed once again, the resulting signal is naturally noisier than the original signal. The second embodiment of the present invention can take advantage of the mechanisms that produce generational loss, by employing those techniques that inherently modify the signal. These mechanisms can be used to naturally produce an output that, for example, has embedded noise which is very close to the masking thresholds depicted in figure 3. Such a result could be obtained by employing a non-linear quantizer in the compression algorithm that is adjusted to more coarsely quantize the individual frequency components of the signal. Thus, this output signal would not be able to undergo a second compression/decompression cycle without the added noise from the second compression cycle being above the masking threshold, and thus being audible in the output signal.

A third approach to implement the second embodiment of the present invention uses the fact that compression algorithms with improved generational qualities often use additional techniques to reduce bit requirements without adding quantization noise. These techniques can provide the basis for further one generation functionality methods. For example, some algorithms, such as the Dolby AC-3 compression algorithm, employ a technique called Huffman encoding in addition to reduced quantization resolution on a frequency band by frequency band basis. Huffman encoding uses the elimination of redundancies in the audio signal over time to reduce data requirements. It decreases the number of bits needed to

described an audio signal by first encoding the audio signal using complete information and then only using differences in this information to describe the audio signal over a defined sequential time interval. Compression algorithms using such a technique have better generational characteristics than those that do not because they can use finer frequency band quantization and still maintain the desired compression ratio. They suffer, however, from having reduced audio data time resolution. The underlying assumption that significant changes in input audio signal characteristics will not take place over the time window used by the Huffman encoding process, can be used by the one generation compression process. One example of such use is the addition by a one generation audio compression process of short duration audio data or noise bursts to its output audio data stream. It is well known in the art that as an audio data sample is reduced in duration it must be of greater amplitude to be perceived by the listener when in the presence of competing sounds. For example, an 8 kHz tone with a duration of 1 millisecond, beginning 2 milliseconds after the initiation of 60 db of Uniform Masking noise, must be 33 dB greater in amplitude as compared to an 8 kHz tone with a duration of 20 milliseconds, beginning 2 milliseconds after the initiation of 60 db of Uniform Masking noise, to be perceived by the human ear. This was reported by H. Fastl in 1976 in his paper 'Temporal masking effects: I. Broad band masker' which appeared in *Acustica*, 35(5), 287-302. Audio data samples which occur randomly in time, or at chosen predetermined time intervals, and are short enough in time duration will therefore not be easily sensed by the listener, but will be detected by an audio compression process attempting to compress the audio signal. Using some of the specific techniques described above, as exemplified in Figures 3 and 4, will further hide the randomly added audio samples from a listener. If this audio compression process employs Huffman encoding, these pulses will asynchronously occur at the time the Huffman encoding process is preparing the data which is used as the reference for subsequent audio difference samples, and cause these subsequent samples to incorrectly represent the audio being compressed. In the case of Dolby AC-3, the Huffman encoding window is 30 milliseconds. This means that the output

compressed audio will be corrupted for 30 milliseconds each time the Huffman reference information is spuriously altered by these embedded short audio noise bursts. This corruption will represent a significant degradation of the decompressed audio signal.

As previously illustrated, some of the specific techniques described add sufficient noise to an audio signal at various frequencies and amplitudes to adversely affect application of a subsequent compression algorithm, but not enough to discernibly affect the quality of the signal without such further compression. A fourth approach applicable to the one generation algorithm of the second embodiment of the current invention shown in Figure 8, uses a different method of accomplishing similar ends. It employs the concept of temporal unmasking. As described above, a usual compression encoding algorithm operates on successive, uniform blocks 529, 531 etc. of digital samples of the signal 527 (Figure 2). If these blocks are not uniform, information defining the timing and number of bytes of data associated with each of these blocks of digital samples must be sent along with the compressed data for use by the compression decoding algorithm in order to reconstruct a replica of the signal 527. It is the alteration of this block timing and block size that can constitute the noise or data added by block 581 in the embodiment of Figure 8, either alone or in combination with some level of spectral alteration.

In one popular compression process, each successive block of audio data includes 256 new time samples as well as the previous 256 time samples. This block of 512 overlapping samples is windowed and the data in this window, which moves in time, is transformed into 256 unique frequency coefficients. In addition, the input signals are analyzed with a high frequency bandpass filter, to detect the presence of transients. This information is used to adjust the block size of the data transformed, restricting quantization noise associated with the transient to within a small temporal region about the transient, avoiding temporal unmasking. The method under consideration utilizes the fact that the changing data block size and/or windowing time position, occurring on compression encode, must be transmitted to

the decompression decoder in order to accurately decompress the encoded audio signal. One method of doing this is through the use of side chain information, although other methods, which embed this information into the compressed audio data stream itself, may be employed. This permits the decoder to accurately synchronize the decode operation with the varying encoded data block size and assure the same block size is employed for decode as was used for encode, thus avoiding temporal unmasking. The present method takes advantage of the fact that this additional side chain information is not included in the decompressed audio data stream and is thus not available to subsequent compression processes. To exploit this circumstance, the present method calls for the one generation compression algorithm under consideration to place transient noise or data at locations in the audio data stream being compressed which is synchronized with the sample block size and sample block timing used during the process of transforming the audio data stream data from the time to the frequency domain. This transient extraneous data is tailored such that the audio data present in the audio signal begin compressed, which occurs immediately before and immediately after the transient, masks the audibility of these transients, so they will not be perceptible to the listener when the audio signal is decompressed. In addition, the one generation compression algorithm under consideration uses a varying sample block size during the process of transforming the data from the time to the frequency domain. Data regarding this varying block size, as well as data regarding where transients were inserted into the audio stream, are transmitted to the decoder by one of several means well known in the art. This data will permit the original audio signal to be decompressed and reproduced with high quality. No transient artifacts would be heard by a listener. However, since block size and transient timing information is not included with the decompressed audio data stream, a subsequent compression process, whether it uses a fixed size window, multiple fixed sized windows or dynamically sized windows to analyzing the spectral and temporal components of the audio signal being compressed, will be unable to select the best window size for transient response, or synchronize the windowing function to the transients that were inserted in the

uncompressed, treated audio stream. This will cause these transients to be temporally unmasked and therefore audible at the output of the second compression decompression cycle. This temporal masking embodiment, as the others, is advantageously implemented in the system described in the above referenced Secure Transmission Patent Application, in order to prevent the consumer from having access to the digital signals from the first compression process before they are converted to PCM or analog signals.

Although the various example implementations of two embodiments of the present invention have been described in the form of fixed algorithms applied to an input audio signal, all of the algorithmic processes described can be adjusted during their application as a function of input audio signal characteristics. The objective of this adjustment is to maximize the difference between the processed audio signal and the processed audio signal after undergoing audio compression. This "adaptive processing", referred to as optimization, can be effectuated by first analyzing the amplitude and timing of the input audio signal's frequency components, as well as the relationship between the audio data present in each channel of the input audio signal, and then using this information to select from a multiple of processing algorithms or to adjust process algorithm parameters and function. Changes to the phase, amplitude and frequency modifications, as well as the character of the spurious data, introduced in the treated audio signal will directly influence both the quality of the uncompressed processed audio signal and the amount the processed audio signal is degraded after compression. The block diagram in Figure 9 depicts anti-compression method 619 which can be used alone to add anti-compression characteristics to uncompressed audio signals or as part of a one generation audio compression codec 619 operates on two channel stereo audio signals and tunes anti-compression processing as a function of input signal characteristics. For a monophonic implementation only blocks 583, 585, 587, 589 and 593 of 619 would be required because the additional blocks shown, 611, 603, 601, 599, 597 and 595, are for second channel relationship analysis and second channel anti-compression processing. For a greater than two channel

implementation, elements of method 619 are replicated to accommodate the processing and relationship analysis required by the additional channels. An instance of blocks 611, 603, 601, 599, 597, and 595 would be required for each additional channel added. In method 619, stereo audio channel number 1 is applied to input line 617 and stereo audio channel number 2 is applied to input line 605. These two audio signals are separated into their individual frequency components by filter bank 583 and filter bank 603 respectively. Although not depicted, the frequency component separation process would normally be digital in nature and require the input signals to first be converted to digital form, if they were not already in digital form when applied. In addition, filter banks 583 and 603 could either be transformed based, as employed by signal modification system 511, or a sub-band based. If a transform based process is employed, a block quantizing step would be required before the frequency component separation step performed by blocks 583 and 603. Method 619 assumes the use of a sub-band based process, so no prior block quantizing step is shown. A sub-band based process uses narrow band time domain filters to continuously partition the input audio signal into its critical frequency bands. The input audio signal is therefore not transformed into its frequency domain representation and thus no block quantizing step is required. The frequency component activity analysis derived by blocks 583 and 603, which corresponds to block spectrum 533 of system 511, is used by blocks 585 and 601 respectively to calculate the masking functions associated with each of the two stereo channels as well as to derive, for example, temporal audio activity, audio signal dynamic range, and audio signal baseline offset. This information is used by spurious signal generator blocks 587 and 599 respectively, often in conjunction with data from signal relationship block 611, to create spurious signals, which are combined with the input stereo signals 617 and 605 by adder blocks 593 and 595, which are output on lines 591 and 621 as anti-compressed treated signals. It is also used by signal modification blocks 589 and 597, also often in conjunction with data from block 611, to alter, but not add to, the signals output on 591 and 621. For example, time related masking curve information from blocks 585 and 601 can be

employed by blocks 587 and 599 to create noise bursts inserted into the output audio signals 591 and 621 that are optimized in both timing and in frequency characteristics, so as to maximally confuse audio compression codecs employing Huffman encoding techniques, as previously described, but which are optimally masked by the audio signal frequency components present so they are not audible to the listener. Also, the frequency and phase relationships between the input audio signals appearing on line 617 and 605, that are derived by the actions of block 611, can be used by audio signal modification blocks 589 and 597 to adaptively shift the relative phase of frequency elements common to both output signals 591 and 621, so as to cause audio compression codecs employing joint stereo encoding techniques to be optimally confused, as previously described, and produce very degraded results. Further, signal relationship data from block 611 can be used by blocks 587 and 599 to add out of phase extraneous signals into each of the output channels, through the use of blocks 593 and 595, that can only be heard if the stereo output signal is compressed with an audio compression codec using absolute value addition techniques, as was also previously described, thus again causing poor results from a subsequent compression/decompression process.

In a typical application of either the first or second embodiment of the present invention, each of multiple incoming audio signals is modified according to a common algorithm. In the event that a computer hacker is able to ascertain that algorithm and then use that information to remove the modifications from an audio signal, the algorithm can be changed by a content provider for subsequent audio signal processing. This would then make it necessary for the hacker to determine the new algorithm each time it is changed. Alternatively, many different algorithms can be alternately used by content providers in order to make the task of removing the modifications from the signal even more difficult. This notion can be taken one step further by using a different algorithm on different parts of the same song. In addition to causing greater challenges for computer hackers in their efforts to compromise the beneficial effects of the audio processing begin disclosed, it will allow a single song to be tailored to the characteristics of multiple audio

compression technologies and thus prevent this processed song from being compressed with quality by a large number of different compression encoder algorithms.

Although the various aspects of the present invention have been described with respect to specific embodiments thereof, it will be understood that the invention is entitled to protection within the full scope of the appended claims.